Resource

WormBase as an Integrated Platform for the *C. elegans* ORFeome

Nansheng Chen,[1,3] Daniel Lawson,[2] Keith Bradnam,[2] Todd W. Harris,[1] and Lincoln D. Stein[1]

[1]*Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA;* [2]*The Wellcome Trust Sanger Institute, Hinxton, CB10 ISA United Kingdom*

The ORFeome project has validated and corrected a large number of predicted gene models in the nematode *C. elegans*, and has provided an enormous resource for proteome-scale studies. To make the resource useful to the research and teaching community, it needs to be integrated with other large-scale data sets, including the *C. elegans* genome, cell lineage, neurological wiring diagram, transcriptome, and gene expression map. This integration is also critical because the ORFeome data sets, like other 'omics' data sets, have significant false-positive and false-negative rates, and comparison to related data is necessary to make confidence judgments in any given data point. WormBase, the central data repository for information about *C. elegans* and related nematodes, provides such a platform for integration. In this report, we will describe how *C. elegans* ORFeome data are deposited in the database, how they are used to correct gene models, how they are integrated and displayed in the context of other data sets at the WormBase Web site, and how WormBase establishes connection with the reagent-based resources at the ORFeome project Web site.

The publication of *C. elegans* ORFeome 1.1 confirmed approximately 4000 gene models that had not previously been confirmed by cDNA or EST data. More than 50% of the 11,984 open reading frames (ORFs) identified by the project showed different intron–exon structures from those given by the ab initio gene prediction program that had been used to predict *C. elegans* genes unsupported by cDNA or EST evidence (Reboul et al. 2001, 2003). As described in other articles in this special issue of *Genome Research*, the ORFeome project provides a collection of reagents that can be used to study gene function or to perform large-scale interactome (genome-wide protein–protein interaction) mapping (Walhout and Vidal 2001; Boone and Andrews 2003; Li et al. 2004). This resource will be used by *C. elegans* researchers for many years to come.

The WormBase database (http://wormbase.org) is an online resource for biological data in *C. elegans* and other nematodes that has been developed and maintained by over 30 scientists and software engineers from four different institutions (http://wormbase.org/about/people.html; Harris et al. 2004). Although the resource is of substantial size (over 8 Gigabytes of data and expanding), it is quite dynamic. It is updated every two weeks, with an average of 120 changes in gene model per release. Data deposited in WormBase come from two major sources: high-throughput projects that generate large data sets, and individual experiments that are curated from the published literature. Examples of large data sets include the genomic sequences of *C. elegans* (The *C. elegans* Sequencing Consortium 1998) and *C. briggsae* (Stein et al. 2003) genome-wide RNAi trials (e.g., Piano et al. 2000; Kamath et al. 2003; Simmer et al. 2003), The Stanford Microarray Database (SMD) microarray expression maps (Kim et al. 2001), deletion alleles from the *C. elegans* knock-out consortium (http://www.celeganskoconsortium.omrf.org/), EST libraries (Kohara and Shin-i 1999), the ORFeome project, the nervous system wiring data (White et al. 1986) and the cell lineage data (Sulston and Horvitz 1977). Data curated from the literature include text descriptions of gene function, reports of new genes based on genetic screens, polymorphisms, and gene mapping information.

In this article, we outline how WormBase accommodates, integrates, and utilizes data from the ORFeome project, how users access and view the data and finally, how users can use WormBase to obtain the ORFeome sequence tags (OSTs) for their own research.

RESULTS

Integration of Data From the ORFeome Project in WormBase

WormBase obtains the ORFeome project primers and OSTs (Reboul et al. 2001, 2003) by retrieving the sequence data from the GenBank EST division using the same processing pipeline as used for nematode EST data. Since the OST sequences are single pass forward and reverse reads from a RT–PCR-derived cDNA, subsequent resequencing of an OST from the same gene will be regarded as a different entity with a different name. The OSTs are mapped to the *C. elegans* genome using BLAT (Kent 2002). The BLAT output is then parsed to assign the single best match as BLAT_OST_BEST (corresponding to "ORFeome sequence tag (best)" in the feature section of the "Batch Sequences" page described below) and any secondary matches as BLAT_OST_OTHER. This assignment is based on the percent identity of the match and the length of the match in relation to the query sequence (OST read length). The mapped OSTs are then used to validate and improve existing gene models at WormBase, as described later in a separate section. The current version of WormBase (release WS122) hosts 19,400 ORFeome project primer pairs, 13,197 pairs of which were successfully amplified. The data set is also available at WorfDB (http://worfdb.dfci.harvard.edu/; Vaglio et al. 2003), a database maintained by the Vidal group.

[3]**Corresponding author.**
E-MAIL chenn@cshl.org; FAX (516) 367-6851.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.2521304.

ORFeome Data for Individual Genes

To describe how the WormBase integration of the ORFeome data can be used, let us consider the hypothetical case of a researcher who is searching for new members of the *odr-10* family, one of the earliest and best studied of the *C. elegans* chemosensory gene superfamily (Sengupta et al. 1996). To search for more instances of this family, the researcher uses the WormBase interface to the BLAST program (Altschul and Gish 1996) to search against the *C. elegans* nucleotide and protein data sets.

At the nucleotide level, one BLAST hit, F10D2.4, matches *odr-10* at an e-value < 10^{-20}. While the gene model inferred from the BLAST hits is similar to the existing F10D2.4 gene model, it is not exactly the same. Specifically, the BLAST hits indicate that the first exon of the F10D2.4 gene model should be extended, exons 5 and 6 should be combined together with the intron between these two exons, and the last exon should be removed from the gene model. One explanation for this is that the WormBase gene model is incorrect; another is that there has been evolutionary change in the gene structure between *odr-10* and F10D2.4. To sort this out, the investigator can turn to the ORFeome data. Starting with the WormBase home page (Harris et al. 2004), the researcher searches for "Any Gene" using the candidate chemosensory receptor gene's name (F10D2.4). This search will lead directly to the gene summary page for F10D2.4 (Fig. 1). This page summarizes key information for the gene of interest, provides links to more specialized pages including the corresponding RNAi page, sequence page, protein page, expression profile page, information on precomputed best BLASTP matches to proteins in other organisms and literature references relevant to the gene. From a list of precomputed BLAST hits on this page (not shown), the user can find close homolog to the F10D2.4 gene, including the *odr-10* gene itself.

The ORFeome data can be found on the gene page in the "Reagents" section, under "ORFeome Project primers & sequences." If the gene has been successfully assayed in the ORFeome project, this section will list three hyper-linked entries, corresponding to the 5′ ORFeome sequence tag, the 3′ ORFeome sequence tag and the ORFeome primers themselves.

Clicking on the 5′ ORFeome sequence tag brings the user to the sequence page (http://www.wormbase.org/db/seq/sequence?name=OSTF176E5_1;class=Sequence; Fig. 2) where the tag sequence is described, a graphical alignment of the tag on the genome, and relevant links to EMBL, GenBank, and the ORFeome project database WorfDB (http://worfdb.dfci.harvard.edu/; Vaglio et al. 2003). The bottom of the page contains the nucleotide sequence of OSTF176E5. In this case, the ORFeome tag validates each of the exons and splice junctions in F10D2.4 (Figs. 2, 3).

Returning to the gene page, the user may click on the third entry in the ORFeome "Project primers & sequences" line in order to navigate to the ORFeome project primers page (http://wormbase.org/db/seq/pcr?name=mv_F10D2.4;class=PCR_product). This page displays the oligonucleotide sequences for both the left and right primers of the ORFeome tag, and again links to the WorfDB database. Unlike the sequence page, this page provides a graphical representation of the ORFeome primers in the context of the F10D2.4 gene model. The reaction conditions for the PCR amplification are listed at the bottom of the page.

Clicking on the graphic in either the Sequence or Orfeome project Primers page leads users to the Genome Browser (http://wormbase.org/db/seq/gbrowse/wormbase?name=V%3A7138817..7141118).

The Genome Browser (Stein et al. 2002) integrates all the information previously displayed in a single configurable view. It shows the structure of the F10D2.4 gene model, the position of the ORFeome project primers and their amplification status, and the structure of the 5′ and 3′ OST tags. In addition, all the other genome annotation information in this region is available, including EST matches, cDNA matches, and the presence of RNAi and expression map data in and around the gene.

WormBase also provides a handy interface between its BLAST tool and the genome browser. To use this interface, the researcher performs a BLAST (TBLASTN) of the *odr-10* protein sequence against the *C. elegans* genome, and then follows the links to the Genome Browser representation of the hit positions. By enabling the ORFeome tracks, the researcher can see exactly how the *odr-10* BLAST results map onto the region around F10D2.4, and compare this to the position of the OSTs (Fig. 3). The BLAST hits are displayed in a track entitled "BlastHSPs_from_Query" in parallel with the gene model and the OST of F10D2.4. This information shows that, on the one hand, the BLAST hits overlap well with the exons of an existing gene model F10D2.4, suggesting that there is likely a valid *odr-10* homologous gene model in this genomic region. On the other hand, the BLAST hits don't match exactly with the gene model. Comparing the gene model with the OST tag confirms that the existing gene model is valid (Fig. 3).

Researchers who wish to look up information on a single primer pair can do so by searching "Primer Pair" from the search menu on the WormBase front page by typing the name of the "Primer Pair" reagent.

Accessing ORFeome Data in Batch Form

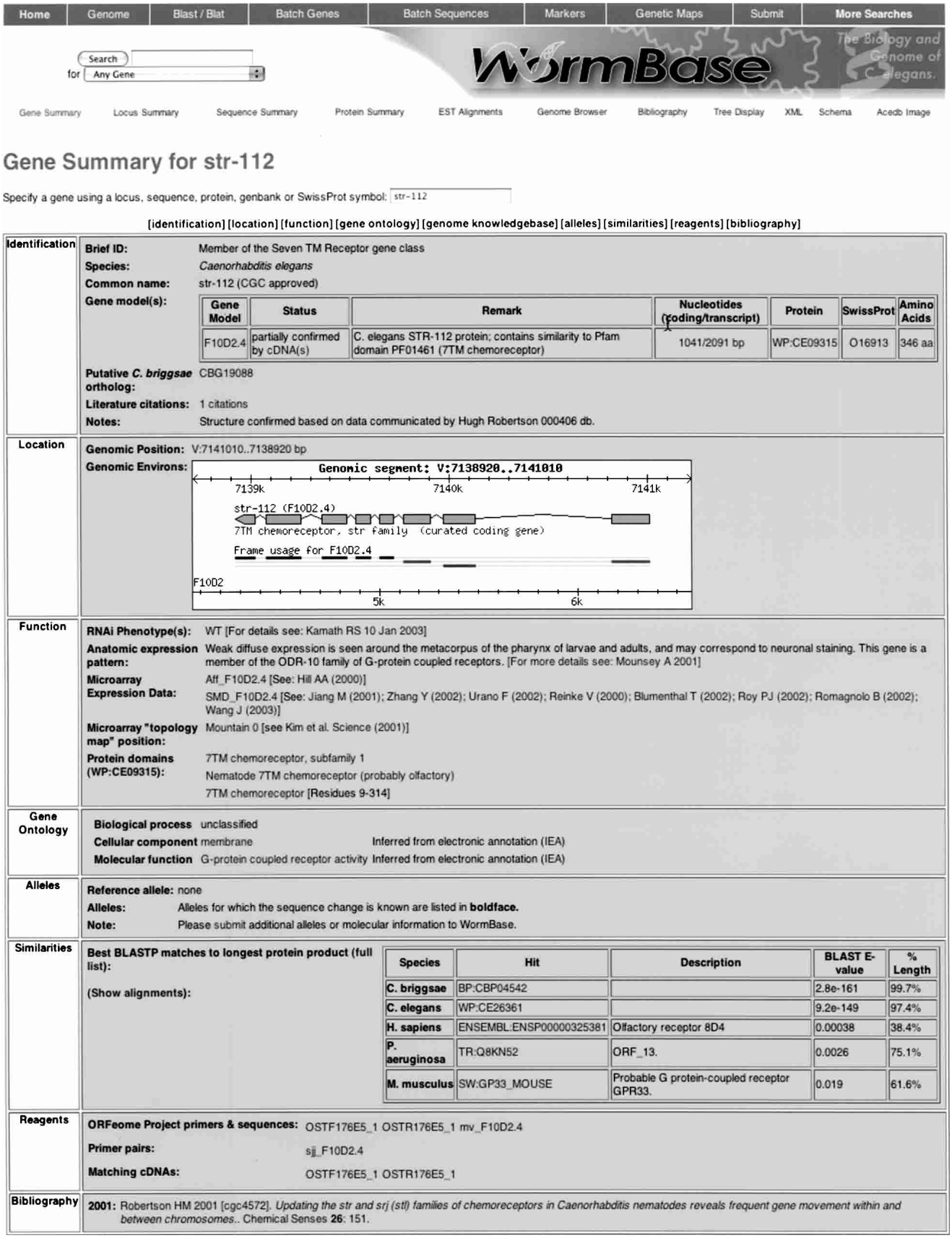
Using *odr-10* as a query to BLAST against the *C. elegans* genome actually yields a large number of significant hits (>300; Robertson et al. 1999; Robertson 2000, 2001). The researcher could examine the BLAST hits one at a time following the procedure described earlier, but there is a better way. Instead, the researcher can obtain ORFeome data in bulk via the "Batch Sequences" page (http://wormbase.org/db/searches/advanced/dumper), accessible from the navigation bar at the top of every WormBase page. To download the alignment of OST sequences to the genes covered by *odr-10,* the researcher inputs the names of these genes in the text field in the left panel of the "Batch Sequences" page. In the middle panel, the researcher selects "ORFeome sequence tags (best)" feature. In the right panel, the researcher only needs to select the output format between plain text and html. The researcher can also select "Save to disk (Plain TEXT)" to download the results. In addition to OST alignments, the "Batch Sequences" page allows researchers to download a large number of sequence features including 5′UTR, 3′UTR, EST, cDNA, gene models, and Operons. The page also allows users to obtain sequence features for *C. briggsae*.

Obtaining ORFeome Reagents for Further Studies

After validating the gene models for the 300+ putative chemosensory receptors, the investigator might want to express these genes to evaluate their function. The ORFeome project offers the necessary resources for this purpose, by providing the OST in an InVitrogen Gateway™ recombinational vector system (Reboul et al. 2003; Li et al. 2004). All of the OSTs are available from the MRC Geneservice, and the appropriate link for ordering the reagents can be found on the ORFeome Primers Page described above.

Complex Queries

In addition to the name-based searches discussed above, WormBase provides users with a rich selection of other ways to query the database (Harris et al. 2004). For example, users can query the



Home | Genome | Blast / Blat | Batch Genes | Batch Sequences | Markers | Genetic Maps | Submit | More Searches

Search for Any Gene

WormBase — The Biology and Genome of C. elegans.

Gene Summary | Locus Summary | Sequence Summary | Protein Summary | EST Alignments | Genome Browser | Bibliography | Tree Display | XML | Schema | Acedb Image

Gene Summary for str-112

Specify a gene using a locus, sequence, protein, genbank or SwissProt symbol: str-112

[identification] [location] [function] [gene ontology] [genome knowledgebase] [alleles] [similarities] [reagents] [bibliography]

Identification

Brief ID: Member of the Seven TM Receptor gene class
Species: *Caenorhabditis elegans*
Common name: str-112 (CGC approved)
Gene model(s):

Gene Model	Status	Remark	Nucleotides (coding/transcript)	Protein	SwissProt	Amino Acids
F10D2.4	partially confirmed by cDNA(s)	C. elegans STR-112 protein; contains similarity to Pfam domain PF01461 (7TM chemoreceptor)	1041/2091 bp	WP:CE09315	O16913	346 aa

Putative *C. briggsae* ortholog: CBG19088
Literature citations: 1 citations
Notes: Structure confirmed based on data communicated by Hugh Robertson 000406 db.

Location

Genomic Position: V:7141010..7138920 bp
Genomic Environs:

Function

RNAi Phenotype(s): WT [For details see: Kamath RS 10 Jan 2003]
Anatomic expression pattern: Weak diffuse expression is seen around the metacorpus of the pharynx of larvae and adults, and may correspond to neuronal staining. This gene is a member of the ODR-10 family of G-protein coupled receptors. [For more details see: Mounsey A 2001]
Microarray Expression Data: Aff_F10D2.4 [See: Hill AA (2000)]
SMD_F10D2.4 [See: Jiang M (2001); Zhang Y (2002); Urano F (2002); Reinke V (2000); Blumenthal T (2002); Roy PJ (2002); Romagnolo B (2002); Wang J (2003)]
Microarray "topology map" position: Mountain 0 [see Kim et al. Science (2001)]
Protein domains (WP:CE09315): 7TM chemoreceptor, subfamily 1
Nematode 7TM chemoreceptor (probably olfactory)
7TM chemoreceptor [Residues 9-314]

Gene Ontology

Biological process unclassified
Cellular component membrane — Inferred from electronic annotation (IEA)
Molecular function G-protein coupled receptor activity — Inferred from electronic annotation (IEA)

Alleles

Reference allele: none
Alleles: Alleles for which the sequence change is known are listed in **boldface.**
Note: Please submit additional alleles or molecular information to WormBase.

Similarities

Best BLASTP matches to longest protein product (full list):
(Show alignments):

Species	Hit	Description	BLAST E-value	% Length
C. briggsae	BP:CBP04542		2.8e-161	99.7%
C. elegans	WP:CE26361		9.2e-149	97.4%
H. sapiens	ENSEMBL:ENSP00000325381	Olfactory receptor 8D4	0.00038	38.4%
P. aeruginosa	TR:Q8KN52	ORF_13.	0.0026	75.1%
M. musculus	SW:GP33_MOUSE	Probable G protein-coupled receptor GPR33.	0.019	61.6%

Reagents

ORFeome Project primers & sequences: OSTF176E5_1 OSTR176E5_1 mv_F10D2.4
Primer pairs: sjj_F10D2.4
Matching cDNAs: OSTF176E5_1 OSTR176E5_1

Bibliography

2001: Robertson HM 2001 [cgc4572]. *Updating the str and srj (stl) families of chemoreceptors in Caenorhabditis nematodes reveals frequent gene movement within and between chromosomes.*. Chemical Senses **26**: 151.

Figure 1 Gene summary page for F10D2.4 (str-112). The gene summary page summarizes key information and serves as a gateway to related information. It is divided into multiple sections with links at the *top* of the table. Pages with ORFeome data, the Genome Browser page, the OST sequence page, and the PCR primer page all have links embedded in the gene summary page.

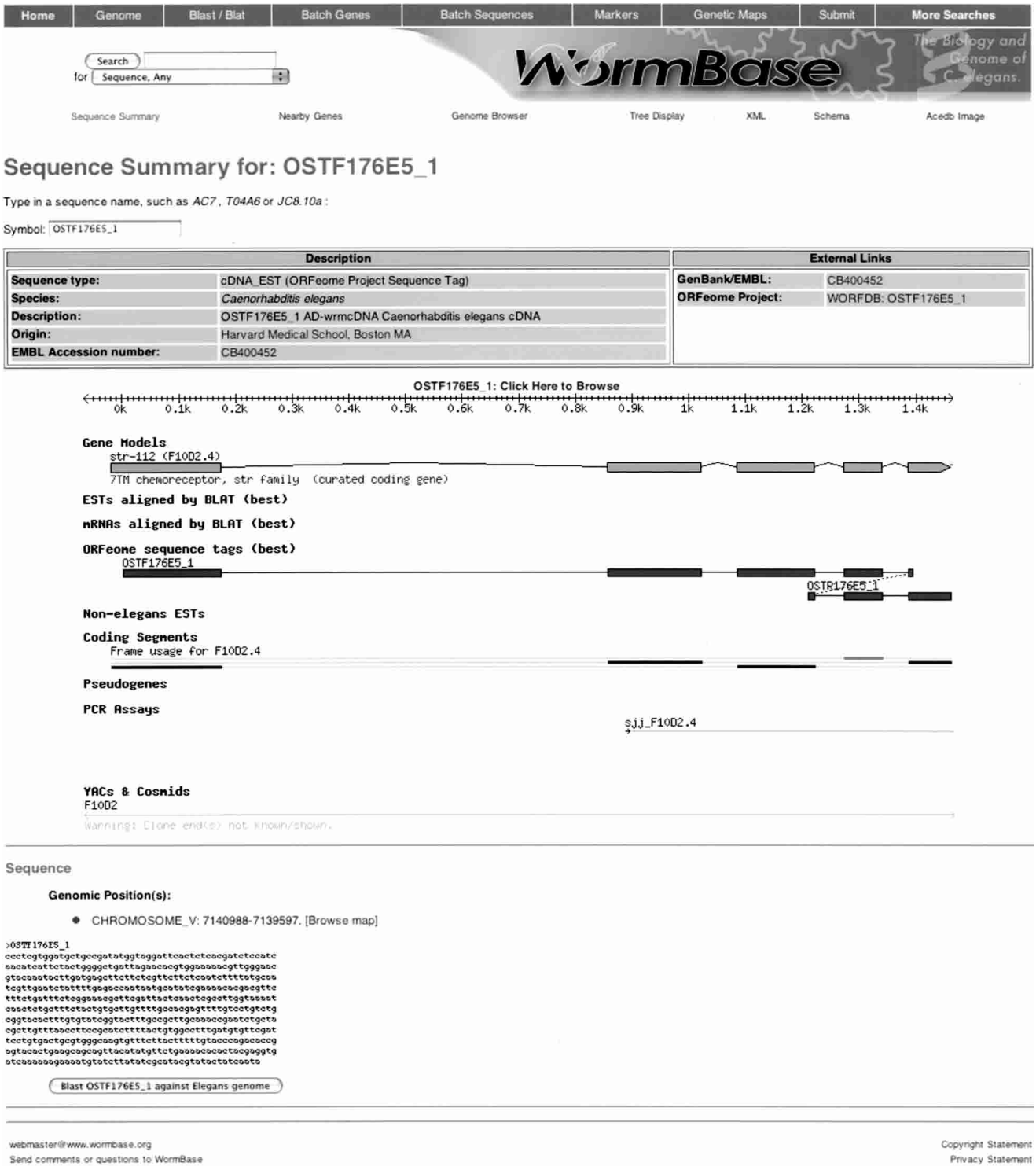


Figure 2 Sequence page for ORFeome sequence tags. The sequence summary page for OSTs contains a description about the OSTs, a graphical representation of the OSTs in the context of the genome, corresponding gene models, and the nucleotide sequence of the OSTs.

database via the "AQL Search" page, which gives users access to ad hoc queries via the Ace Query Language (AQL). This page provides a large number of examples for users to follow and learn. With some practice, users should be able to do fairly complex queries. For example, if a user want to retrieve the spliced DNA sequences of all genes on chromosome X that have ORFeome models and have been assayed in RNAi trials, the user can type the following query into the text field on the "AQL Search" page:

```
select a from a in class CDS
where a→Sequence→Interpolated_map_position like "X"
and  a→Corresponding_pcr_product→amplified = 1
and  exists_tag  a→RNAi_result
```

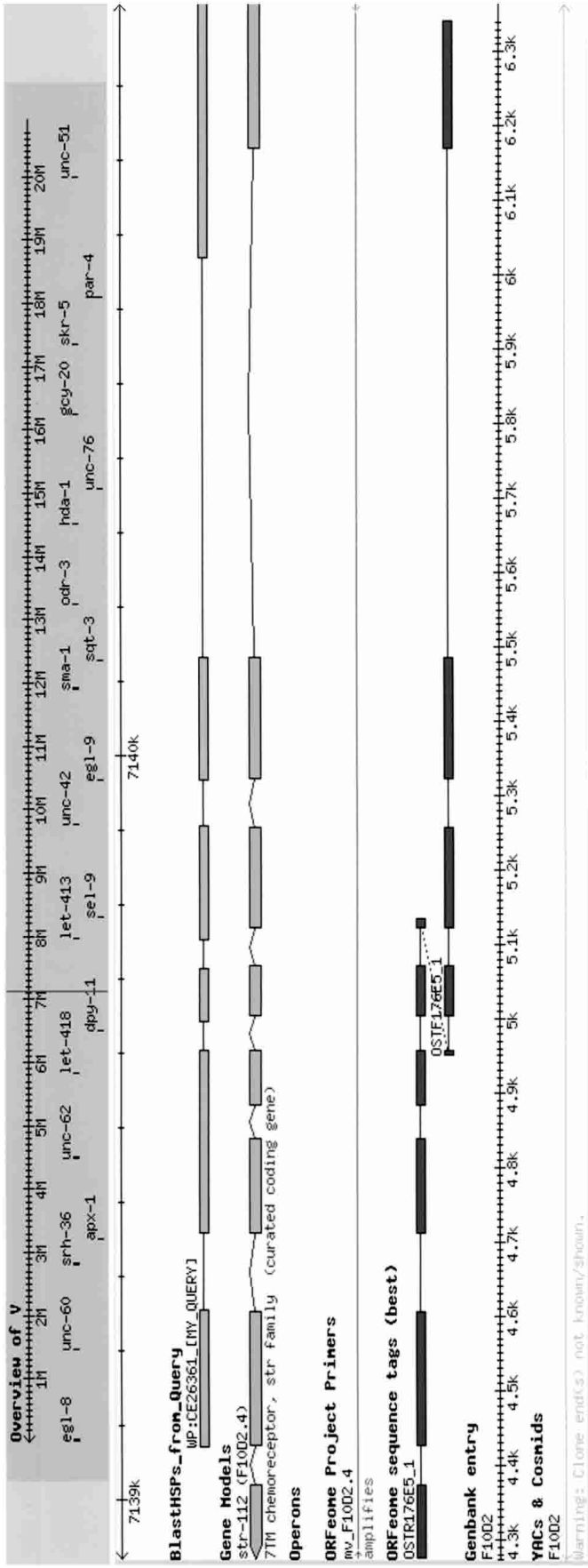


Figure 3 BLAST hits displayed in the genome browser. The BLAST/BLAT page displays links to the genome browser with the BLAST results shown in relation to the gene models and other corresponding features including the OSTs. This figure shows the genome browser view of BLAST hits with *odr-10* as query against the *C. elegans* genome. The 'BlastHSPs_from_Query' track shows the BLAST hits. Other tracks shown include Gene Models, ORFeome project Primers, and ORFeome sequence tags (best).

This query returns a list of 1524 genes, which can then be queried further. This query can be easily modified to retrieve genes on other chromosomes and with other properties. WormBase is scheduled to provide a wizard-interface to this query feature in the fourth quarter of 2004 based on the BioMart technology (Kasprzyk et al. 2004).

For more advanced users with some programming experiences, WormBase has set up a data mining server for remote access. Details about how to connect to the database and query the database using Perl modules are included in the data-mining page (http://wormbase.org/about/linking.html#mining), accessible from the WormBase front page.

Using ORFeome Data to Improve *C. elegans* Gene Models

In addition to providing researchers with access to the ORFeome data set, WormBase is actively using the ORFeome information to improve the canonical set of *C. elegans* gene models. WormBase gene models are examined and updated following a two-step protocol. First, WormBase curators mark up putative gene prediction corrections based on cases in which OSTs and gene models don't agree. This step is performed systematically, and utilizes in-house software that compares the OSTs to the gene models and flags disagreements. Second, the problematic cases are examined by hand and updates are made based on the judgment of the curator. Most of the judgment calls involve deciding whether a gap in an aligned OST corresponds to a proper intron. Criteria for accepting a gap as an intron include: Does the gap look like an intron with canonical donor/acceptor regions (GT..AG)? Is the gap biologically long enough to be an intron (>30 bp)? If these criteria are met then the gap is declared to correspond to confirmed intron feature between the two blocks of similarity. This process is repeated for all OST reads.

During the curation process, curators visualize the alignment of OSTs to the genome and to existing gene models using ACeDB (Stein and Thierry-Mieg 1998). In cases where the alignment will cause a change to one or more of the boundaries of the gene's coding region (CDS), the updated CDS is written back to ACeDB, and the change is propagated automatically into the gene's corresponding protein translation. As with gene models and other genomic features, the OSTs and updated gene models are also written out into Gene Feature Format (GFF, originally proposed by Durbin and Haussler, http://www.sanger.ac.uk/Software/GFF), a flat file format that is very amenable to processing via scripts written in Perl (http://www.perl.org). This allows the OST annotation data to be processed via a variety of quality control scripts.

At the current time, approximately 1300 gene models (or about 6.5% of the all of the gene models in the whole genome, excluding isoforms) have been improved based on the OSTs from the *C. elegans* ORFeome projects. In addition, based on the OST reads WormBase has increased the number of fully confirmed genes (those in which each splice junction has been verified and every single base of the gene has been confirmed) by approximately 600 (15% increase compared to the number of genes known at the time). The number of partially confirmed coding regions increased by 3200 (37%). Information on the changes to the modified genes can be obtained from the WormBase gene summary pages by following the curators' "Notes" field. Altogether, 8545 genes have supporting data in the ORFeome data set (Reboul et al. 2003).

DISCUSSION

The whole ORFeome project data set has been integrated into WormBase, where it can be searched, browsed, and downloaded in the context of other large-scale and curated *C. elegans* data sets. As the ORFeome project makes new releases, the data set contained within WormBase will continue to be updated. In addition, we will continue to improve *C. elegans* gene models as new ORFeome data (and other relevant data) become available. The data model used in WormBase will allow us to accommodate ORFeome-style data from other nematode genomes should they become available in the future.

With the availability of the *C. elegans* OSTs in the InVitrogen Gateway recombinational vector system, we foresee that increasing amounts of large-scale functional assay data for *C. elegans* will be produced by the research community. A prime example of this is the genome-wide *C. elegans* interactome data published earlier this year (Li et al. 2004). We are committed to integrating this and future ORFeome-derived data sets into WormBase, where they can be queried, browsed and downloaded by the community.

ACKNOWLEDGMENTS

WormBase is supported by grant P41-HG02223 from the U.S. National Human Genome Research Institute and the British Medical Research Council. WormBase is a collaborative effort of the authors and the following scientists and software engineers: Payan Canaran and Fiona Cunningham from Cold Spring Harbor Laboratory; Igor Antoshechkin, Carol Bastiani, Juancarlos Chan, Wen Chen, Eimear Kenny, Ranjana Kishore, Raymond Lee, Hans-Michael Muller, Cecilia Nakamura, Andrei Petcherski, Erich Schwarz, Paul Sternberg, Kimberly Van Auken, and Daniel Wang from California Institute of Technology; Tamberlyn Bieri, Darin Blasiar, Phil Ozersky, and John Spieth from Washington University at St. Louis; Chao-Kung Chen, Paul Davis, Richard Durbin, and Anthony Rogers from the Wellcome Trust Sanger Institute. We want to thank Dr. Doreen Ware for critical reading of the manuscript.

REFERENCES

Altschul, S.F. and Gish, W. 1996. Local alignment statistics. *Methods Enzymol.* **266:** 460–480.

Boone, C. and Andrews, B. 2003. ORFeomics: Correcting the wiggle in worm genes. *Nat. Genet.* **34:** 8–9.

The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* **282:** 2012–2018.

Harris, T.W., Chen, N., Cunningham, F., Tello-Ruiz, M., Antoshechkin, I., Bastiani, C., Bieri, T., Blasiar, D., Bradnam, K., Chan, J., et al. 2004. WormBase: a multi-species resource for nematode biology and genomics. *Nucleic Acids Res.* **32:** D411–D417.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., et al. 2003. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421:** 231–237.

Kasprzyk, A., Keefe, D., Smedley, D., London, D., Spooner, W., Melsopp, C., Hammond, M., Rocca-Serra, P., Cox, T., and Birney, E. 2004. EnsMart: a generic system for fast and flexible access to biological data. *Genome Res.* **14:** 160–169.

Kent, W.J. 2002. BLAT—the BLAST-like alignment tool. *Genome Res.* **12:** 656–664.

Kim, S.K., Lund, J., Kiraly, M., Duke, K., Jiang, M., Stuart, J.M., Eizinger, A., Wylie, B.N., and Davidson, G.S. 2001. A gene expression map for *Caenorhabditis elegans*. *Science* **293:** 2087–2092.

Kohara, Y. and Shin-i, T. 1999. NEXTDB: the nematode expression pattern map database. In *Proceedings of the International C. elegans* Meeting, pp. 776. University of Wisconsin, Madison, WI.

Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.O., Han, J.D., Chesneau, A., Hao, T., et al. 2004. A map of the interactome network of the metazoan *C. elegans*. *Science* **303:** 540–543.

Piano, F., Schetter, A.J., Mangone, M., Stein, L., and Kemphues, K.J. 2000. RNAi analysis of genes expressed in the ovary of *Caenorhabditis elegans*. *Curr. Biol.* **10:** 1619–1622.

Reboul, J., Vaglio, P., Tzellas, N., Thierry-Mieg, N., Moore, T., Jackson, C., Shin-i, T., Kohara, Y., Thierry-Mieg, D., Thierry-Mieg, J., et al. 2001. Open-reading-frame sequence tags (OSTs) support the existence of at least 17,300 genes in *C. elegans*. *Nat. Genet.* **27:** 332–336.

Reboul, J., Vaglio, P., Rual, J.F., Lamesch, P., Martinez, M., Armstrong, C.M., Li, S., Jacotot, L., Bertin, N., Janky, R., et al. 2003. *C. elegans* ORFeome version 1.1: Experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nat. Genet.* **34:** 35–41.

Robertson, H.M. 2000. The large *srh* family of chemoreceptor genes in *Caenorhabditis* nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. *Genome Res.* **10:** 192–203.

———. 2001. Updating the str and srj (stl) families of chemoreceptors in *Caenorhabditis* nematodes reveals frequent gene movement within and between chromosomes. *Chem. Senses* **26:** 151–159.

Robertson, H.M., Martos, R., Sears, C.R., Todres, E.Z., Walden, K.K., and Nardi, J.B. 1999. Diversity of odourant binding proteins revealed by an expressed sequence tag project on male Manduca sexta moth antennae. *Insect Mol. Biol.* **8:** 501–518.

Sengupta, P., Chou, J.H., and Bargmann, C.I. 1996. odr-10 encodes a seven transmembrane domain olfactory receptor required for responses to the odorant diacetyl. *Cell* **84:** 899–909.

Simmer, F., Moorman, C., Van Der Linden, A.M., Kuijk, E., Van Den Berghe, P.V., Kamath, R., Fraser, A.G., Ahringer, J., and Plasterk, R.H. 2003. Genome-wide RNAi of *C. elegans* using the hypersensitive rrf-3 strain reveals novel gene functions. *PLoS Biol* **1:** E12.

Stein, L.D. and Thierry-Mieg, J. 1998. Scriptable access to the *Caenorhabditis elegans* genome sequence and other ACEDB databases. *Genome Res.* **8:** 1308–1315.

Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A., et al. 2002. The generic genome browser: A building block for a model organism system database. *Genome Res.* **12:** 1599–1610.

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., et al. 2003. The genome sequence of *Caenorhabditis briggsae*: A platform for comparative genomics. *PLoS Biol.* **1:** E45.

Sulston, J.E. and Horvitz, H.R. 1977. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* **56:** 110–156.

Vaglio, P., Lamesch, P., Reboul, J., Rual, J.F., Martinez, M., Hill, D., and Vidal, M. 2003. WorfDB: the *Caenorhabditis elegans* ORFeome database. *Nucleic Acids Res.* **31:** 237–240.

Walhout, A.J. and Vidal, M. 2001. Protein interaction maps for model organisms. *Nat. Rev. Mol. Cell. Biol.* **2:** 55–62.

White, J.G., Southgate, E., Thomson, J.N., and Brenner, F.R.S. 1986. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **314:** 1–340.

WEB SITE REFERENCES

http://wormbase.org; The model organism database for the biology and genomics of *Caenorhabditis elegans* and *Caenorhabditis briggsae*.

http://www.celeganskoconsortium.omrf.org/; *Caenorhabditis elegans* knockout project; Oklahoma Medical Research Foundation, USA, University of British Columbia, Canada, and The Genome Sciences Center, BC Cancer Research Center, Canada.

http://worfdb.dfci.harvard.edu/; *Caenorhabditis elegans* ORFeome project led by Marc Vidal Laboratory.

http://www.sanger.ac.uk/Software/GFF; General Feature Format.

http://www.perl.org; The official Web site for the programming language Perl.

Received February 27, 2004; accepted in revised form April 16, 2004.

WormBase as an Integrated Platform for the *C. elegans* ORFeome

Nansheng Chen, Daniel Lawson, Keith Bradnam, et al.

Genome Res. 2004 14: 2155-2161
Access the most recent version at doi:10.1101/gr.2521304